Theoretical Computer Science 410 (2009) 5261–5272

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Expander properties and the cover time of random intersection graphs ☆

S. Nikoletseas, C. Raptopoulos *, P.G. Spirakis

University of Patras, 26500 Patras, Greece
Computer Technology Institute, P.O. Box 1122, 26110 Patras, Greece

ARTICLE INFO

Keywords:
Random intersection graphs
Random walks
Cover time
Mixing time
Expansion

ABSTRACT

We investigate important combinatorial and algorithmic properties of $G_{n,m,p}$ random intersection graphs. In particular, we prove that with high probability (a) random intersection graphs are expanders, (b) random walks on such graphs are "rapidly mixing" (in particular they mix in logarithmic time) and (c) the cover time of random walks on such graphs is optimal (i.e. it is $\Theta(n \log n)$). All results are proved for p very close to the connectivity threshold and for the interesting, non-trivial range where random intersection graphs differ from classical $G_{n,p}$ random graphs.

1. Introduction

Random graphs are interesting combinatorial objects that were introduced by Erdös and Rényi and still attract a great deal of research in the communities of Theoretical Computer Science, Algorithms, Graph Theory and Discrete Mathematics. This continuing interest is due to the fact that, besides their mathematical beauty, such graphs are very important, since they can model interactions and faults in networks and also serve as typical inputs for an average case analysis of algorithms.

There exist various models of random graphs. The most famous is the $G_{n,p}$ random graph, a sample space whose points are graphs produced by randomly sampling the edges of a graph on n vertices independently, with the same probability p. Other models have also been investigated quite a lot: $G_{n,r}$ (the "random regular graphs", produced by randomly and equiprobably sampling a graph from all regular graphs of n vertices and vertex degree r) and $G_{n,M}$ (produced by randomly and equiprobably selecting an element of the class of graphs on n vertices having M edges). For an excellent survey of these models, see [2,4].

In this work we study important properties (expansion properties and the cover time) of a relatively recent model of random graphs, namely the random intersection graphs model introduced by Karoński, Sheinerman and Singer-Cohen [15,26]. Also, Godehardt and Jaworski [12] considered similar models. In $G_{n,m,p}$, to each of the n vertices of the graph, a random subset of a universal set of m elements is assigned, by independently choosing elements with the same probability p. Two vertices u, v are then adjacent in the $G_{n,m,p}$ graph if and only if their assigned sets of elements have at least one element in common.

Importance and motivation. First of all, we note that (as proved in [16]) any graph is a random intersection graph. Thus, the $G_{n,m,p}$ model is very general. Furthermore, for some ranges of the parameters m, p ($m = n^{\alpha}, \alpha > 6$) the spaces $G_{n,m,p}$ and

☆ This work was partially supported by the IST Programme of the European Union under contract number IST-2005-15964 (AEOLUS) and by the Programme PENED under contract number 03ED568, co-funded 75% by European Union – European Social Fund (ESF), 25% by Greek Government – Ministry of Development – General Secretariat of Research and Technology (GSRT), and by Private Sector, under Measure 8.3 of O.P. Competitiveness – 3rd Community Support Framework (CSF). Also by the EU within the 6th Framework Programme under contract 001907 DELIS.

* Corresponding author at: University of Patras, 26500 Patras, Greece.
E-mail addresses: nikole@cti.gr (S. Nikoletseas), raptopox@ceid.upatras.gr (C. Raptopoulos), spirakis@cti.gr (P.G. Spirakis).

0304-3975/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2009.08.028

$G_{n,\hat{p}}$, for $\hat{p} = 1 - \left(1 - \frac{p^2}{(1-p)^2 + np(1-p) + \binom{n}{2}p^2}\right)^m$, are asymptotically equivalent (as proved by Fill, Sheinerman and Singer-Cohen [11], showing that in this range, the total variation distance between the graph random variables has limit 0 as n goes to ∞).

Second, random intersection graphs may model real-life applications more accurately (compared to the $G_{n,\hat{p}}$ Bernoulli random graphs case). In fact, there are practical situations where each communication agent (e.g. a wireless node) gets access only to some ports (statistically) out of a possible set of communication ports. When another agent also selects a communication port, then a communication link is implicitly established and this gives rise to communication graphs that look like random intersection graphs. Even epidemiological phenomena (like spread of disease) tend to be more accurately captured by this "proximity-sensitive" random intersection graphs model. Other applications may include oblivious resource sharing in a distributed setting, interactions of mobile agents traversing the web etc.

Regarding the properties we study, we believe that their importance is evident. So we just mention the fact that expander graphs are basic building blocks in optimal network design. Also, at a combinatorial/algorithmic level, it is well known that random walks whose second largest eigenvalue is sufficiently less than 1 are "rapidly mixing", i.e. they get close (in terms of the variation distance) to the steady state distribution after only a polylogarithmic (in the number of vertices/states) number of steps (see e.g. [25]); this has important algorithmic applications e.g. in efficient random generation and counting of combinatorial objects. Finally, the cover time of a graph is one of its most important combinatorial measures which also captures practical quantities like the expected communication time in a network of mobile entities, infection times in security applications etc.

Related work. Random intersection graphs have recently attracted a growing research interest. The question of how close $G_{n,m,p}$ and $G_{n,\hat{p}}$ are for various values of m, p has been studied by Fill, Sheinerman and Singer-Cohen in [11]. In [18], new models of random intersection graphs have been proposed, along with an investigation of both the existence and efficient finding of close to optimal independent sets. The authors of [10] find thresholds (that are optimal up to a constant factor) for the appearance of hamilton cycles in random intersection graphs. The efficient construction of hamilton cycles is studied in [23]. Also, by using a sieve method, Stark [27] gives exact formulae for the degree distribution of an arbitrary fixed vertex of $G_{n,m,p}$ for a quite wide range of the parameters of the model. In [20], the authors use a coupling technique to bound the second eigenvalue of random walks on instances of symmetric random intersection graphs $G_{n,n,p}$ (i.e. in random intersection graphs with $m = n$), when p is near the connectivity threshold. The upper bound proved holds for almost every instance of the symmetric random intersection graphs model. We should note that in this paper we deal with the case $m = n^\alpha$, for $\alpha < 1$, which is very different from the symmetric case, as in the first case each label is selected by a large number of vertices (which allows for much tighter concentration bounds that help in the analysis).

In [14] the author proves that with high probability (whp) the cover time (that is the *expected time* to visit all the vertices of the graph) of a simple random walk on a Bernoulli random graph $G_{n,\hat{p}}$ is quite close to optimal when $\hat{p} = \Omega\left(\frac{\ln n}{n}\right)$. Also, he proves that by further increasing the value of $\hat{p}$, the same bound that holds for the cover time holds whp for the actual time needed for the random walk on $G_{n,\hat{p}}$ to visit all the vertices of the graph. His results are improved by Cooper and Frieze in [5], who prove that when $\hat{p} = \frac{c \log n}{n}$, $c > 1$, the cover time of $G_{n,\hat{p}}$ is asymptotic to $c \log\left(\frac{c}{c-1}\right) n \log n$.

Geometric proximity between randomly placed objects is also nicely captured by the model of random geometric graphs (see e.g. [7,8,22]) and important variations (like random scaled sector graphs, [9]). In [3], the cover time of random geometric graphs near the connectivity threshold is found almost optimal, by showing that the effective resistance of the graph is small. Other extensions of random graph models (such as random regular graphs) and several important combinatorial properties (connectivity, expansion, existence of a giant connected component) are performed in [17,21].

Our contribution. As proved in [11], the spaces $G_{n,m,p}$ and $G_{n,\hat{p}}$ are equivalent when $m = n^\alpha$, with $\alpha > 6$, in the sense that their total variation distance tends to 0 as n goes to ∞. Also, the authors in [23] show that, when $\alpha > 1$, for any monotone increasing property there is a direct relation (including a multiplicative constant) of the corresponding thresholds of the property in the two spaces. So, it is very important to investigate what is happening when $\alpha \leq 1$ where the two spaces are statistically different. In this paper, we study the regime $\alpha < 1$. In particular

(a) We first prove that $G_{n,m,p}$ random intersection graphs are c-expanders (i.e. every set S of at most $n/2$ vertices is connected to at least $c|S|$ other vertices outside S) with high probability. This is shown for $p = \frac{\ln n + g(n)}{m}$, where $g(n) \to \infty$ arbitrarily slowly, i.e. p is *just* above the connectivity threshold.[1] Note that [20] has no equivalent results to this one.

(b) We then show that random walks on the vertices of random intersection graphs are whp rapidly mixing (in particular, the mixing time is logarithmic on n). This is shown for p very close to the connectivity threshold τ_c of $G_{n,m,p}$, with $m = n^\alpha$, $\alpha < 1$. We interestingly note that the c-expansion property shown in (a) cannot ensure "small" rapid mixing. For example imagine the following graph pointed out to us by Noga Alon [1]: two cliques of size $n/2$ each, connected by a perfect matching of their vertices is a c-expander but has mixing time $\Omega(n)$. To get our result on the mixing time we had to prove an upper bound on the second eigenvalue of $G_{n,m,p}$, that holds with high probability, through coupling arguments of the original Markov Chain describing the random walk and another Markov Chain on an associated random bipartite graph whose conductance properties we show to be appropriate. The attentive reader can easily understand

[1] The connectivity threshold for $\alpha \leq 1$ is proved to be $\tau_c = \frac{\ln n}{m}$ in [26].

that although the general technique used to prove the results of Section 4 is similar to the technique used in [20], the proofs are quite different. More specifically, in the case of $G_{n,m,p}$, with $m = n^\alpha, \alpha < 1$, the concentration results of Section 2 (and especially the first part of Lemma 1) can be used to give an elegant proof of Lemma 4 (which cannot be applied in the symmetric case considered in [20]).

(c) Finally, we show that the cover time of such graphs (in the interesting, non-trivial range mentioned above and for p close to the connectivity threshold) is whp $\Theta(n \ln n)$, i.e. optimal up to multiplicative constants. To get this result we had to prove a technically involved intermediate result relating the probability that our random walk on $G_{n,m,p}$ has not visited a vertex v by time t with the degree of v. Note that [20] has no equivalent results to this one. Also, to prove the results of Section 6, one needs to prove an extra preliminary result (namely Lemma 3) that does not appear in [20].

A preliminary version of this paper appeared in MFCS 2007 ([19]).

2. Notation, definitions and properties of $G_{n,m,p}$

Let $Bin(n, p)$ denote the Binomial distribution with parameters n and p. We first formally define the *random intersection graphs model*.

Definition 1 (*Random Intersection Graph*). Consider a universe $\mathcal{M} = \{1, 2, \ldots, m\}$ of elements and a set of vertices $V(G) = \{v_1, v_2, \ldots, v_n\}$. If we assign independently to each vertex $v_j, j = 1, 2, \ldots, n$, a subset S_{v_j} of $\mathcal{M}$ choosing each element $i \in \mathcal{M}$ independently with probability p and put an edge between two vertices v_{j_1}, v_{j_2} if and only if $S_{v_{j_1}} \cap S_{v_{j_2}} \neq \emptyset$, then the resulting graph is an instance of the random intersection graph $G_{n,m,p}$. In this model we also denote by L_l the set of vertices that have chosen label $l \in M$. The degree of $v \in V(G)$ will be denoted by $d_G(v)$. Also, the set of edges of $G_{n,m,p}$ will be denoted by $e(G)$.

Consider now the bipartite graph with vertex set $V(G) \cup \mathcal{M}$ and edge set $\{(v_j, i) : i \in S_{v_j}\} = \{(v_j, i) : v_j \in L_i\}$. We will refer to this graph as the *bipartite random graph* $B_{n,m,p}$ *associated to* $G_{n,m,p}$.

In this section we assume that $m = n^\alpha$, for some $\alpha < 1$. This is the interesting regime where $G_{n,m,p}$ differs from $G_{n,\hat{p}}$ (see also "Our Contribution" in the previous section). Let $\tau_c \stackrel{def}{=} \frac{\ln n}{m}$ be the connectivity threshold for $G_{n,m,p}$ in that case. Also we assume that $p = 4\tau_c$. We prove that

Lemma 1. *The following hold with high probability in $G_{n,m,p}$ when $\alpha < 1$ and $p = 4\frac{\ln n}{m}$*

(a) *For every label $l \in \mathcal{M}$ we have $(1-\epsilon)np \le |L_l| \le (1+\epsilon)np$ for any $\epsilon \in [n^{-(1-\alpha)/2}, 1)$.*
(b) *For every vertex $v \in V$ we have $|S_v| \in (1 \pm \sqrt{4/5})4 \ln n$.*

Proof. (a) By the definition of the model we have that $|L_l|$ follows $Bin(n, p)$. So, by using Chernoff bounds and Boole's inequality we get that for any constant $\epsilon \in [n^{-(1-\alpha)/2}, 1)$

$$Pr(\exists l \in \mathcal{M} : ||L_l| - np| \ge \epsilon np) \le m \exp\left\{-\frac{\epsilon^2 np}{3}\right\} = o(1)$$

since $np = 4n^{1-\alpha} \ln n$.

(b) By the definition of the model we have that $|S_v|$ follows $Bin(m, p)$. So by using Chernoff bounds and Boole's inequality

$$Pr(\exists v \in V : ||S_v| - mp| \ge \sqrt{4/5}mp) \le n \exp\left\{-\frac{16}{15}\ln n\right\} = o(1). \quad \square$$

Note that Lemma 1 implies that the minimum degree in $G_{n,m,p}$ when $\alpha < 1$ and p just above the connectivity is whp at least $\Omega(n^{1-\alpha} \ln n)$. In fact, we prove the following

Lemma 2. *The following hold with high probability in $G_{n,m,p}$ when $\alpha < 1$ and $p = 4\frac{\ln n}{m}$*

(a) *The degree of any vertex $v \in V$ satisfies $d_G(v) \in (1 \pm n^{-\epsilon'})4|S_v|n^{1-\alpha} \ln n$ for any constant $\epsilon' \in (0, 1-\alpha)$.*
(b) *The number of edges of the $G_{n,m,p}$ graph satisfies $|e(G)| \in (1 \pm \epsilon'')8n^{2-\alpha} \ln^2 n$, for any small constant $\epsilon'' > 0$.*
(c) *There are no vertices $x \neq y \in V(G)$ such that $|S_x \cap S_y| \ge \lceil \frac{3}{\alpha} \rceil$.*

Proof. (a) Let L_{l_1,l_2} be the set of vertices that have chosen both labels l_1 and l_2. Then $|L_{l_1,l_2}|$ follows $Bin(n, p^2)$ and $E|L_{l_1,l_2}| = 4n\frac{\ln^2 n}{m^2} = 4n^{1-2\alpha} \ln^2 n = \mu$. Then by using Chernoff bounds, for $\delta = n^{\alpha-b}$ where $b \in (0, \alpha)$ is a constant (bounded away from 0 and α), we have

$$Pr(\exists l_1, l_2 \in \mathcal{M} : |L_{l_1,l_2}| > (1+\delta)\mu) \le m^2 \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu = o(1).$$

But by Lemma 1, any vertex v chooses $O(\ln n)$ labels, which means that, for any $\epsilon' \in (0, b)$ (bounded away from 0 and b), every label it has chosen has at least $(1 - n^{-\epsilon'})4n^{1-\alpha} \ln n$ vertices that belong to none of the other labels in S_v. This proves (a).

(b) Since $|e(G)| = \frac{1}{2}\sum_{v\in V} d_G(v)$ and $|S_v|$ follows $Bin(m, p)$ (so that $E|S_v| = mp = 4\ln n$), this is an application of (a) and the Strong Law of Large Numbers.

(c) Set $i = \lceil 3/\alpha \rceil$. A very crude inequality suffices to give

$$Pr(\exists x \neq y \in V(G) : |S_x \cap S_y| \geq i) \leq n^2 \binom{m}{i} p^{2i} = o(1)$$

for $p = 4\frac{\ln n}{m}$. □

Let now $D(k)$ be the number of vertices $v \in V(G)$ that have $|S_v| = k$, i.e. $D(k) = |\{v \in V : |S_v| = k\}|$. Consider also the following partition of the set $\{0, 1, \ldots, m\}$.

$$M_2 = \{k \in (1 \pm \sqrt{4/5}) \ln n : E[D(k)] > \ln n\}$$
$$M_1 = \{k \in (1 \pm \sqrt{4/5}) \ln n : E[D(k)] \leq \ln n\}$$
$$M_0 = \{\{0, 1, \ldots, m\} \backslash \{M_2 \cup M_1\}\}.$$

We can then prove the following lemma that will be useful for upper bounding the cover time.

Lemma 3. *For the $G_{n,m,p}$ with $\alpha < 1$ and $p = 4\tau_c$ the following hold with high probability*

1. *For every $k \in M_0$, $D(k) = 0$*
2. *For every $k \in M_1$, $D(k) \leq \ln^3 n$ and*
3. *For every $k \in M_2$, $D(k) \leq 2E[D(k)]$.*

Proof. (1) This follows immediately from (b) of Lemma 1.

(2) Note that by using Markov's inequality and Boole's inequality we have

$$Pr(\exists k \in M_1 : D(k) > \ln^3 n) \leq \sum_{k\in M_1} \frac{E[D(k)]}{\ln^3 n} = O\left(\frac{1}{\ln n}\right)$$

where for the final inequality we used that $E[D(k)] \leq \ln n$ and $|M_1| = O(\ln n)$.

(3) By the definition of the model we have that $D(k)$ follows $Bin(n, Pr(|S_v| = k))$, where v is some fixed vertex and $Pr(|S_v| = k) = \binom{m}{k} p^k (1-p)^{m-k}$. So, by using Chernoff bounds we get

$$Pr(\exists k \in M_2 : D(k) > 2E[D(k)]) \leq O(\ln n)\mathrm{e}^{-\Theta(\ln n)} = o(1).$$

This completes the proof of the Lemma. □

3. Expansion properties of random intersection graphs

We first give the following definition:

Definition 2 (*c-expanders*). Let c be a positive constant. A graph $G = (V(G), E(G))$ is a c-expander if every set $S \subseteq V(G)$ of at most $n/2$ vertices is connected to at least $c|S|$ vertices outside S.

In this section we assume that $p = \frac{\ln n + g(n)}{m}$, that is, p is *just* above the connectivity threshold τ_c. In the following, let $S_X = \bigcup_{v\in X} S_v$, for $X \subseteq V$ and let $L_Y = \bigcup_{l\in Y} L_l$, for $Y \subseteq \mathcal{M}$. We prove the following:

Theorem 1. *Assume that $m = n^\alpha$, $\alpha < 1$ and $p = \frac{\ln n + g(n)}{m}$, for some function $g(n) \to \infty$ (arbitrarily slowly). With high probability $G_{n,m,p}$ is a c-expander, for some constant $c > 0$.*

Proof. By the definition of c-expanders, we need to show that every set $X \subseteq V(G)$ with at most $\frac{n}{2}$ vertices has at least $c|X|$ neighbouring vertices outside X. We distinguish three cases.

Case I: $\frac{n}{2} \geq |X| \geq \frac{n}{\ln n}$. We show that whp every such set X has many labels. Let $\epsilon > 0$ a small constant. Then

$$\begin{aligned} &Pr\left(\exists X : \frac{n}{\ln n} \leq |X| \leq \frac{n}{2}, |S_X| \leq (1-\epsilon)m\right) \\ &\leq \sum_{x=\frac{n}{\ln n}}^{n} \binom{n}{x} \sum_{i=0}^{(1-\epsilon)m} \binom{m}{i} (1-(1-p)^x)^i (1-p)^{x(m-i)} \\ &\leq \sum_{x=\frac{n}{\ln n}}^{n} \binom{n}{x} \left((1-p)^{xm} + \sum_{i=1}^{(1-\epsilon)m} \left(\frac{me}{i}\right)^i \mathrm{e}^{-pxm+pxi}\right) \end{aligned}$$

$$\leq \sum_{x=\frac{n}{\ln n}}^{n} \binom{n}{x} \left(e^{-pxm} + \sum_{i=1}^{(1-\epsilon)m} e^{i+i\ln m-i\ln i-pxm+pxi} \right)$$

$$\leq \sum_{x=\frac{n}{\ln n}}^{n} \left(\frac{ne}{x}\right)^{x} e^{-\epsilon pxm+O(m\ln m)}$$

$$\leq \sum_{x=\frac{n}{\ln n}}^{n} e^{x\ln\ln n-\epsilon x\ln n+O(x)} = o(1). \tag{1}$$

So, whp every set $X \subseteq V$ of $G_{n,m,p}$ with size at least $\frac{n}{\ln n}$ spans at least $(1-\epsilon)m$ labels, where ϵ is an arbitrarily small positive constant.

Now we show that whp every set of labels Y of size at least $(1-\epsilon)m$ contains at least $\left(\frac{1}{2}+c_1\right)n$ vertices, for any constant $c_1 \in \left(0, \frac{1}{2}\right)$ (or equivalently, there are at least $\left(\frac{1}{2}+c_1\right)n$ vertices each of which contains at least one label in Y).

$$Pr\left(\exists Y : (1-\epsilon)m \leq |Y| \leq m, |L_Y| \leq \left(\frac{1}{2}+c_1\right)n\right)$$

$$\leq \sum_{y=(1-\epsilon)m}^{m} \binom{m}{y} \sum_{j=0}^{(1/2+c_1)n} \binom{n}{j} (1-(1-p)^y)^j (1-p)^{y(n-j)}$$

$$\leq \sum_{y=(1-\epsilon)m}^{m} \binom{m}{y} \left((1-p)^{yn} + \sum_{j=1}^{(1/2+c_1)n} \left(\frac{ne}{j}\right)^{j} e^{-pyn+pyj} \right)$$

$$\leq \sum_{y=(1-\epsilon)m}^{m} \left(\frac{me}{y}\right)^{y} e^{-(1/2-c_1)pyn+O(n)}$$

$$\leq \sum_{y=(1-\epsilon)m}^{m} e^{-(1/2-c_1)p(1-\epsilon)mn+O(n)} = o(1). \tag{2}$$

This ends the proof for the first case, since inequalities (1) and (2) imply that whp *every* set X of size at least $\frac{n}{\ln n}$ has at least $c_1|X|$ neighbours outside X.

Case II: $n^{1-\alpha} \leq |X| < \frac{n}{\ln n}$. We first show that whp every such set X has at least $\frac{|X|m}{n}$ labels.

$$Pr\left(\exists X : n^{1-\alpha} \leq |X| < \frac{n}{\ln n}, |S_X| \leq \frac{|X|m}{n}\right)$$

$$\leq \sum_{x=n^{1-\alpha}}^{\frac{n}{\ln n}} \binom{n}{x} \sum_{i=0}^{xm/n} \binom{m}{i} (1-(1-p)^x)^i (1-p)^{x(m-i)}$$

$$\leq \sum_{x=n^{1-\alpha}}^{\frac{n}{\ln n}} \binom{n}{x} \left((1-p)^{xm} + \sum_{i=1}^{xm/n} \left(\frac{me}{i}\right)^{i} e^{-pxm+pxi} \right)$$

$$\leq \sum_{x=n^{1-\alpha}}^{\frac{n}{\ln n}} \left(\frac{ne}{x}\right)^{x} e^{-pxm+\frac{px^2m}{n}+\frac{xm}{n}\ln m+O(xm/n)}$$

$$\leq \sum_{x=n^{1-\alpha}}^{\frac{n}{\ln n}} e^{-x\ln x+o(x\ln x)} = o(1). \tag{3}$$

Thus, whp every set $X \subseteq V$ of $G_{n,m,p}$ with size $n^{1-\alpha} \leq |X| \leq \frac{n}{\ln n}$ spans at least $\frac{|X|m}{n}$ labels.

Now we show that whp every nonempty set of labels Y of size at most $\frac{m}{\ln n}$ contains at least $(1+c_2)\frac{|Y|n}{m}$ vertices, for any positive constant c_2 (or equivalently, there are at least $(1+c_2)\frac{|Y|n}{m}$ vertices each of which contain at least one label in Y).

$$Pr\left(\exists Y : 1 \leq |Y| \leq \frac{m}{\ln n}, |L_Y| \leq (1+c_2)\frac{|Y|n}{m}\right)$$

$$\le \sum_{y=1}^{\frac{m}{\ln m}} \binom{m}{y} \sum_{j=0}^{(1+c_2)\frac{yn}{m}} \binom{n}{j} (1-(1-p)^y)^j (1-p)^{y(n-j)}$$

$$\le \sum_{y=1}^{\frac{m}{\ln m}} \binom{m}{y} \left((1-p)^{yn} + \sum_{j=1}^{(1+c_2)\frac{yn}{m}} \left(\frac{ne}{j}\right)^j \mathrm{e}^{-pyn+pyj} \right)$$

$$\le \sum_{y=1}^{\frac{m}{\ln m}} \left(\frac{me}{y}\right)^y \mathrm{e}^{-(1-\alpha)pyn+O(yn/m)}$$

$$\le \sum_{y=1}^{\frac{m}{\ln m}} \mathrm{e}^{-(1-\alpha)pyn+O(yn/m)} = o(1). \tag{4}$$

This ends the proof for the second case, since inequalities (3) and (4) imply that whp *every* set X of size $n^{1-\alpha} \le |X| < \frac{n}{\ln n}$ has at least $c_2|X|$ neighbours outside X.

Case III: $|X| < n^{1-\alpha}$. This case is almost trivial. Indeed, note that for any label $l \in \mathcal{M}$, $|L_l|$ follows the binomial distribution with parameters n, p. Hence, mimicking the proof of Lemma 1(a), for any constant $b > 0$ as small as possible, we have

$$Pr(\exists l \in \mathcal{M} : ||L_l| - np| \ge bnp) \le m \exp\left\{-\frac{b^2 np}{3}\right\} = o(1)$$

because $np \ge n^{1-\alpha} \ln n$. Since p is above the connectivity threshold, every vertex in X must have chosen at least one label whp, which means that it will be connected to at least $(1-b)n^{1-\alpha} \ln n = \omega(|X|)$ other vertices.

Setting $c = \min\{c_1, c_2\}$, we conclude that every $X \subseteq V(G)$ with at most $\frac{n}{2}$ vertices has at least $c|X|$ neighbouring vertices outside X and so $G_{n,m,p}$ is almost surely a c-expander. Also note that since c_2 can be set equal to c_1, the expansion factor c equals c_1 and so it can actually be arbitrarily close to $\frac{1}{2}$ (say 0, 49999), which is optimal up to additive constants. □

4. Bounds for the second eigenvalue and the mixing time

In this section we give an upper bound on the second eigenvalue (i.e. the eigenvalue with the largest absolute value less than 1) of $G_{n,m,p}$, with $\alpha < 1$ and $p = 4\tau_c$, that holds for almost every instance. This will imply a logarithmic mixing time.

Let $\tilde{W}$ be a Markov Chain on state space V (i.e. the vertices of $G_{n,m,p}$) and transition matrix given by

$$\tilde{P}(x,y) = \begin{cases} \sum_{l \in S_x \cap S_y} \frac{1}{|S_x| \cdot |L_l|} & \text{if } S_x \cap S_y \ne \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

Note that this Markov Chain comes from observing the simple random walk on the $B_{n,m,p}$ graph associated with $G_{n,m,p}$ every two steps. This means that $\tilde{W}$ is reversible and we can easily verify that its stationary distribution is given by

$$\tilde{\pi}(x) = \frac{|S_x|}{\sum_{y \in V} |S_y|}, \quad \text{for every } x \in V.$$

Now let W denote the random walk on $G_{n,m,p}$ and let P denote its transition probability matrix, that is

$$P(x,y) = \begin{cases} \frac{1}{d_G(x)} & \text{if } S_x \cap S_y \ne \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

It is known that W is reversible and its stationary distribution is given by $\pi(x) = \frac{d_G(x)}{2|e(G)|}$, for every $x \in V$.

Notice now that $P(x,y) > 0 \Leftrightarrow \tilde{P}(x,y) > 0$. By using Theorem 2.1 of [6], we can show that if λ_1 (respectively $\tilde{\lambda}_1$) is the second largest eigenvalue of P (respectively $\tilde{P}$), then

$$\lambda_1 \le 1 - \frac{\beta}{A}(1-\tilde{\lambda}_1) \tag{5}$$

where β is such that $\tilde{\pi}(x) \ge \beta\pi(x)$, for all $x \in V$, and

$$A = \max_{(x,y):P(x,y)>0} \left\{ \frac{\tilde{\pi}(x)\tilde{P}(x,y)}{\pi(x)P(x,y))} \right\}.^{2}$$

Because of (a) of Lemma 2 and the fact that $|e(G)| = \sum_{y\in V} d_G(v)$, there exist two positive constants $\beta < 1 < \beta'$, such that $\beta\pi(x) \le \tilde{\pi}(x) \le \beta'\pi(x)$, for all $x \in V$ (These two constants can be quite close to 1 because of Lemma 2). Also, by (a) and (c) of Lemma 2, for any (x, y) such that $P(x, y) > 0$, we have

$$\frac{\tilde{P}(x,y)}{P(x,y)} = \frac{\sum_{l\in S_x\cap S_y} \frac{1}{|S_x|\cdot|L_l|}}{\frac{1}{d_G(x)}} \le (1+\gamma')|S_x \cap S_y| \le \gamma$$

for some positive constants γ', γ. This means that A is upper bounded by some constant. Thus, we have established that $\lambda_1 \le 1 - \zeta_1(1 - \tilde{\lambda}_1)$, for $\zeta_1 = \frac{\beta}{\beta'\gamma}$.[3]

We now show that $\tilde{\lambda}_1$ is whp bounded away from 1 by some constant (which will mean that λ_1 is also bounded away from 1).

Let $\hat{W}$ denote the random walk on the $B_{n,m,p}$ bipartite graph that is associated to $G_{n,m,p}$. Let also $\hat{P}$ denote its transition probability matrix and let $\hat{\lambda}_i, i = 0, \ldots, m+n-1$, its eigenvalues and $\hat{x}_i, i = 0, \ldots, m+n-1$, their corresponding eigenvectors. Note that

$$\hat{P}^2 = \begin{bmatrix} \tilde{P} & \emptyset \\ \emptyset & Q \end{bmatrix}$$

where Q is some transition matrix. Note that $\hat{P}^2$ has the same eigenvectors as $\hat{P}$ and its eigenvalues are $\hat{\lambda}_i^2, i = 0, \ldots, m+n-1$. Moreover, it is easy to verify that $\tilde{\lambda}_1$ must be an eigenvalue of $\hat{P}^2$. However, the second largest eigenvalue of $\hat{P}^2$ is equal to 1 and so we cannot use it to bound $\tilde{\lambda}_1$ (the latter being strictly less than 1 whp since $\tilde{W}$ is ergodic). So, we use the Markov chain $\hat{W}'$ with the same state space as $\hat{W}$ and transition probability matrix $(\hat{P} + I)/2$, where I is the identity matrix. It is evident that $\hat{W}'$ is ergodic, has stationary state probability $\hat{\pi}$ and has (nonnegative) eigenvalues $\frac{\hat{\lambda}_i+1}{2}, i = 0, \ldots, m+n-1$. By the above discussion, if the second largest eigenvalue of $\hat{W}'$, denoted by $\hat{\lambda}'_1$, is bounded away from 1, then so is $\tilde{\lambda_1}$. In order to bound $\hat{\lambda}'_1$, we use the notion of *conductance* $\Phi_{\hat{W}'}$ of the walk $\hat{W}'$ that is defined as follows:

Definition 3. Consider the bipartite random graph $B_{n,m,p}$ that is associated to $G_{n,m,p}$. The vertex set of $B_{n,m,p}$ is $V(B) = V(G) \cup \mathcal{M}$. For every $x \in V(B)$, let $d_B(x)$ be the degree of x in B. For any $S \subseteq V(B)$, let $e_B(S : \overline{S})$ be the set of edges of S with one end in S and the other in $\overline{S} = V(B)\backslash S$, let $d_B(S) = \sum_{v\in S} d_G(v)$ and $\hat{\pi}(S) = \sum_{v\in S} \hat{\pi}(v)$. Then, the *conductance* $\Phi_{\hat{W}'}$ of the Markov Chain $\hat{W}'$ is

$$\Phi_{\hat{W}'} = \min_{\hat{\pi}(S)\le 1/2} \frac{\sum_{x\in S, y\notin S} \hat{\pi}(x)\frac{1}{2}\hat{P}_{x,y}}{\sum_{x\in S} \hat{\pi}(x)} = \frac{1}{2} \min_{\hat{\pi}(S)\le 1/2} \frac{|e_B(S:\overline{S})|}{d_B(S)}.$$

We now prove the following

Lemma 4. *With high probability, the conductance of the Markov Chain $\hat{W}'$ on $B_{n,m,p}$ satisfies $\Phi_{\hat{W}'} \ge \zeta_2$, where ζ_2 is some positive constant.*

---

[2] The original theorem is as follows: For each pair $x \ne y$ with $\tilde{P}(x, y) > 0$, fix a sequence of steps $x_0 = x, x_1, x_2, \ldots, x_k = y$ with $P(x_i, x_{i+1}) > 0$. This sequence of steps is called a *path* γ_{xy} of length $|\gamma_{xy}| = k$. Let $\mathcal{E} = \{(x, y) : P(x, y) > 0)\}$, $\tilde{\mathcal{E}} = \{(x, y) : \tilde{P}(x, y) > 0)\}$ and $\tilde{\mathcal{E}}(z, w) = \{(x, y) \in \tilde{\mathcal{E}} : (z, w) \in \gamma_{xy}\}$, where $(z, w) \in \mathcal{E}$. Then

$$\lambda_1 \le 1 - \frac{\beta}{A}(1 - \tilde{\lambda}_1)$$

where β is such that $\tilde{\pi}(x) \ge \beta\pi(x)$, for all $x \in V$, and

$$A = \max_{(z,w)\in\mathcal{E}} \left\{ \frac{1}{\pi(x)P(x,y))} \sum_{\tilde{\mathcal{E}}(z,w)} |\gamma_{xy}|\tilde{\pi}(x)\tilde{P}(x,y) \right\}.$$

In our case we have taken $\gamma_{x,y} = \{x_0 = x, x_1 = y\}$ for every $(x, y) \in \tilde{\mathcal{E}}$ which simplifies our formula.

[3] In what follows we will assume without loss of generality (w.o.l.g.) that λ_1 is the eigenvalue of P that has the second greatest absolute value (hence $\lambda_1 > 0$). Indeed, if it was not the case and $1 = \lambda_0 > \lambda_1 \ge \lambda_2 \cdots \ge \lambda_{n-1} > -1$ were the n eigenvalues of P, then λ_{n-1} would be the second greatest eigenvalue, in absolute value (hence $\lambda_{n-1} < 0$). Then we would apply Theorem 2.2 of [6] (which is quite similar to Theorem 2.1 that we used in the text) for bounding λ_{n-1} from below, and the rest of the proofs would be similar.

Proof. All we have to do is to bound $\min_{\hat{\pi}(S)\leq 1/2} \frac{|e_B(S:\bar{S})|}{d_B(S)}$ away from 0. In the following $c_1, c_2, \ldots$ are all constants. Let $S = V_1 \cup M_1$, where $V_1 \subseteq V(G)$ and $M_2 \subseteq \mathcal{M}$. The cases $|V_1| = 0$ or $|M_1| = 0$ are trivial so we do not consider them. We have

$$\hat{\pi}(S) = \hat{\pi}(V_1) + \hat{\pi}(M_1) = \sum_{v\in V_1} \frac{|S_v|}{2\sum_{j\in V(G)}|S_j|} + \sum_{l\in M_1} \frac{|L_l|}{2\sum_{i\in\mathcal{M}}|L_i|}$$
$$\geq \frac{|V_1|(1-\sqrt{4/5})}{2n(1+2\epsilon'')} + \frac{|M_1|(1-o(1))}{2m}$$

where ϵ'' is the constant of (b) of Lemma 2 and for the inequality we used Lemma 1. Since ϵ'' can be as small as possible, we get $\frac{1-\sqrt{4/5}}{1+2\epsilon''} \geq \frac{1}{5}$ and so, bearing in mind that the maximum in the definition of conductance is taken over all S with $\hat{\pi}(S) \leq 1/2$, we have that

$$|V_1| \leq 5n\left(1-(1-o(1))\frac{|M_1|}{m}\right). \tag{6}$$

Notice now that, because of Lemma 1, the number of edges coming out of V_1 is whp $|e_B(V_1 : \mathcal{M})| \in |V_1|(1\pm\sqrt{4/5})4\ln n$. Similarly, the number of edges coming out of M_1 is whp $|e_B(M_1 : V(G))| \in |M_1|(1\pm\epsilon)4n^{1-\alpha}\ln n$. It is then obvious that if $\left(1+\frac{1}{2}\right)|e_B(M_1 : V(G))| \leq |e_B(V_1 : \mathcal{M})|$ or $|e_B(V_1 : \mathcal{M})| \leq \left(1-\frac{1}{2}\right)|e_B(M_1 : V(G))|$, then we are done, since there will be a constant fraction of the number of edges in $e_B(S : S)$ that "leave" S. So we only have to deal with the case $|e_B(V_1 : \mathcal{M})| \in \left(1\pm\frac{1}{2}\right)|e_B(M_1 : V(G))|$, or more simply, $|V_1| \in \left(\frac{1}{10}n^{1-\alpha}|M_1|, 10n^{1-\alpha}|M_1|\right)$. Combining this with the restriction (6) we see that we cannot have $|V_1| \geq \left(1-\frac{1}{40}\right)n$ or $|M_1| \geq \left(1-\frac{1}{40}\right)m$ and at the same time having $|V_1| \in \left(\frac{1}{10}n^{1-\alpha}|M_1|, 10n^{1-\alpha}|M_1|\right)$, because then we would have $\hat{\pi}(S) > 1/2$.

Note now that $|e_B(S : S)|$ follows $Bin(|V_1||M_1|, p)$. So, by using Chernoff bounds and Boole's inequality we can show that, for any constant $c_1 \geq \mathrm{e}^2$,

$$Pr\left(\exists V_1, M_1 : |V_1| = \Theta(n^{1-\alpha}|M_1|), |e_B(S:S)| > \left(1+c_1\frac{m}{|M_1|}\right)p|V_1||M_1|\right)$$
$$\leq \sum_{j=1}^{m}\sum_{i:i=\Theta(n^{1-\alpha}j)}\binom{m}{j}\binom{n}{i}\left(\frac{\mathrm{e}^{c_1\frac{m}{j}}}{\left(1+c_1\frac{m}{j}\right)^{\left(1+c_1\frac{m}{j}\right)}}\right)^{pij}$$
$$\leq \sum_{j=1}^{m}\sum_{i:i=\Theta(n^{1-\alpha}j)}\exp\left\{j\ln m + i\ln n + c_1 4i\ln n - c_1 4i\ln n\ln\left(1+c_1\frac{m}{j}\right)\right\} = o(1).$$

So, whp there is no $S = V_1 \cup M_1$ such that $|V_1| = \Theta(n^{1-\alpha}|M_1|)$ and the edges inside S surpass their mean value by more than $c_1\frac{m}{|M_1|}$ times.

Similarly, note that $|e_B(S : \bar{S})|$ follows $Bin(|V_1|(m-|M_1|) + |M_1|(n-|V_1|), p)$. Let c_2 be a *fixed* large enough positive constant (say $c_2 \geq 70$). We consider two cases:

Case I: Assume that $|V_1|(m-|M_1|)+|M_1|(n-|V_1|) \geq c_2\left(1+c_1\frac{m}{|M_1|}\right)|V_1||M_1|$. So, by using Chernoff bounds and Boole's inequality we can show that, for any constant $c_3 \in \left(\frac{1}{\sqrt{c_2}}, 1\right]$,

$$Pr\left(\exists V_1, M_1 : |V_1| = \Theta(n^{1-\alpha}|M_1|), |e_B(S:\bar{S})| \leq (1-c_3)c_2\left(1+c_1\frac{m}{|M_1|}\right)p|V_1||M_1|\right)$$
$$\leq \sum_{j=1}^{m}\sum_{i:i=\Theta(n^{1-\alpha}j)}\exp\left\{j\ln m + i\ln n - \frac{c_3^2}{2}c_2\left(1+c_1\frac{m}{j}\right)\frac{4\ln n}{m}ij\right\} = o(1).$$

Case II: Assume that $|V_1|(m-|M_1|) + |M_1|(n-|V_1|) < c_6\left(1+c_5\frac{m}{|M_1|}\right)|V_1||M_1|$. Because of the restriction (6), the only way for this to happen is to have $|V_1| = \Theta(n)$ and $|M_1| = \Theta(m)$. But then, because of the restriction posed in the sizes of V_1, M_1, this means that both $|e_B(S : S)|$ and $|e_B(S : \bar{S})|$ have $\Theta(mn)$ candidate edges (which is quite large). By then using Chernoff bounds and Boole's inequality, we can see that the probability that there exist such sets V_1, M_1 that have $|e_B(S : \bar{S})| \leq (1-c_4)p(|V_1|m + |M_1|n - 2|V_1||M_1|)$ or $|e_B(S : S)| \geq (1+c_5)p|V_1||M_1|$ for some constants $c_4, c_5 \in (0, 1]$ (for example we can take $c_4 = c_5 = 0.5$), is at most

$$\sum_{j=\Theta(m)}\sum_{i=\Theta(n)}\binom{m}{j}\binom{n}{i}\mathrm{e}^{-\Theta(n\ln n)} = o(1).$$

We have then proved that $\frac{|e_B(S:\bar{S})|}{|e_B(S:S)|}$ is whp lower bounded by some constant, for any S with $\pi(S) \leq 1/2$. This gives the desired bound on the conductance of $\hat{W}'$. $\square$

By a result of [13,24], we know that $\hat{\lambda}_1' \leq 1 - \frac{\Phi^2_{\hat{W}'}}{2}$ and so $\hat{\lambda}_1'$ is (upper) bounded away from 1. By the above discussion, we have proved the following

Theorem 2. *With high probability, the second largest eigenvalue of the random walk on $G_{n,m,p}$, with $m = n^\alpha$, $\alpha < 1$ and $p = 4\tau_c$, satisfies $\lambda_1 \leq \zeta$, where $\zeta \in (0, 1)$ is a constant that is bounded away from 1.*

Such a bound on λ_1 implies (as shown in Proposition 1 of [25]) a logarithmic mixing time. Thus we get

Theorem 3. *With high probability, there exists a sufficiently large constant $K > 0$ such that if $\tau_0^{(G)} = K \log n$, then for all $v, u \in V(G)$ and any $t \geq \tau_0^{(G)}$,*

$$|P^{(t)}(u, v) - \pi(v)| = O(n^{-3})$$

where $P^{(t)}$ denotes the t-step transition matrix of the random walk W on $G_{n,m,p}$, with $m = n^\alpha$, $\alpha < 1$ and $p = 4\tau_c$. We will refer to $\tau_0^{(G)}$ as the mixing time *of $G_{n,m,p}$.*

5. A useful lemma

In order to give bounds to the cover time of $G_{n,m,p}$, for $m = n^\alpha$, $\alpha < 1$ and p four times the connectivity threshold, we first prove a lemma that bounds the probability that the random walk on $G_{n,m,p}$ has not visited a vertex v by time t by a function of the degree of v. Before presenting the lemma we give some notation.

Let G be an instance of the random intersection graphs model and let $H(v) = G - \{v\}$. We will sometimes write H instead of $H(v)$ when v is clear from the context. Let $\tau_0^{(H)}$ denote the mixing time of H, namely the time needed for the random walk on H to get closer than $O(n^{-3})$ to its steady state distribution (see also definition of $\tau_0^{(G)}$ in Theorem 3). Note that because of the definition of H and by Lemma 1, the removal of v from G does not affect its mixing time[4] and so $\tau_0^{(G)} \sim \tau_0^{(H)} \leq \tau_0 \stackrel{def}{=} \Theta(\log n)$ for any v whp. We will denote by $\mathcal{W}_{u,H}$ the random walk on H that starts at vertex $u \in V(H)$. Let also $\mathcal{W}_{u,H}(t)$ be the random walk generated by the first t steps. For $u \neq v \in V$, let $A_t(v)$ be the event that $\mathcal{W}_{u,G}(t)$ has not visited v.

Lemma 5. *Let G be an instance of $G_{n,m,p}$, with $m = n^\alpha$, $\alpha < 1$ and $p = 4\frac{\ln n}{m}$, that satisfies Lemma 1 and has $\tau_0^{(H)} \leq \tau_0 = \Theta(\log n)$ for every $H = H(v)$ (note that almost every instance of $G_{n,m,p}$ in this range satisfies these requirements). Let δ_v be the minimum degree of the neighbours of $v \in V(G)$. Then, for every $v \in V(G)$,*

$$Pr(A_t(v)) \leq \left(1 - \left(\frac{\delta_v - 1}{\delta_v} - o(1)\right)\frac{d_G(v)}{2|e(G)|}\right)^{t-\tau_0} Pr(A_{\tau_0}(v)).$$

Proof. Fix $w \neq v$ and $y \in N_G(v)$. Note that all these neighbours of v belong to $V(H(v))$. Let $\mathcal{W}_k(y)$ denote the set of walks of length τ_0 in $H(v)$ which start at w, finish at y and leave a vertex of $N_G(v)$ exactly k times. Let $\mathcal{W}_k = \bigcup_y \mathcal{W}_k(y)$ and let $W = (w_0, w_1, \ldots, w_{\tau_0}) \in \mathcal{W}_k(y)$. In order to compare the walk W that happens on $H(v)$ with the same walk (i.e. a walk that passes through the same sequence of vertices) that happens in G, use the quantity

$$\rho_W = \frac{Pr(X_{w,G}(s) = w_s, s = 0, 1, \ldots, \tau_0)}{Pr(X_{w,H}(s) = w_s, s = 0, 1, \ldots, \tau_0)}$$

where $X_{w,G}(s)$ (respectively $X_{w,H}(s)$) denotes the vertex that the walk on G (respectively $H(v)$) occupies at time s. Note that the only difference in transition probabilities between the walk W in $H(v)$ and the same walk in G is when W "leaves" a vertex of $N_G(v)$. Since W leaves such a vertex exactly k times and the minimum vertex degree of some $y \in N_G(v)$ is δ_v, we have that

$$1 \geq \rho_W \geq \left(\frac{\delta_v - 1}{\delta_v}\right)^k.$$

Define the event $\mathcal{E} = \{X_{w,G}(\tau) \neq v, 0 \leq \tau \leq \tau_0\}$ and note that (by definition of $H(v)$) a walk in $\mathcal{W}_k$ never passes through v. Let also $p_{k,y}$ denote the probability that $W_{w,H}(\tau_0)$ leaves a vertex of $N_G(v)$ exactly k times, given that $X_{w,H}(\tau_0) = y$, i.e.

$$p_{k,y} = \frac{Pr(W_{w,H}(\tau_0) \in \mathcal{W}_k(y))}{Pr(X_{w,H}(\tau_0) = y)}.$$

---

4 In fact the same analysis of Section 4 can be applied unchanged to $H(v)$.

So we have

$$\frac{Pr(X_{w,G}(\tau_0)=y|\mathcal{E})}{Pr(X_{w,H}(\tau_0)=y)} = \frac{\sum_{k\geq 0}\sum_{W\in\mathcal{W}_k(y)} Pr(W_{w,G}(\tau_0)=W|\mathcal{E})}{Pr(X_{w,H}(\tau_0)=y)}$$
$$= \sum_{k\geq 0}\sum_{W\in\mathcal{W}_k(y)} \rho_W \frac{Pr(W_{w,H}(\tau_0)=W)}{Pr(X_{w,H}(\tau_0)=y)} = \sum_{k\geq 0} p_{k,y}\rho_W$$
$$\geq \sum_{k\geq 0} p_{k,y}\left(\frac{\delta_v-1}{\delta_v}\right)^k \geq \left(\frac{\delta_v-1}{\delta_v}\right)^{\tau_0}.$$

By Lemma 1 we see that whp $\delta_v \geq n^{1-\alpha}\ln n$, so for $\tau_0 = o(\delta_v)$ we get

$$\frac{Pr(X_{w,G}(\tau_0)=y|\mathcal{E})}{Pr(X_{w,H}(\tau_0)=y)} \geq \left(\frac{\delta_v-1}{\delta_v}\right) - O\left(\frac{\tau_0}{\delta_v}\right). \tag{7}$$

Now if we take w as $X_{u,G}(t-\tau_0-1)$ and assuming $A_{t-\tau_0-1}$, inequality (7) implies

$$\frac{Pr(X_{u,G}(t-1)=y|A_{t-1}(v))}{Pr(X_{w,H}(\tau_0)=y)} = \frac{Pr(X_{w,G}(\tau_0)=y|\mathcal{E})}{Pr(X_{w,H}(\tau_0)=y)} \geq \frac{\delta_v-1}{\delta_v} - O\left(\frac{\tau_0}{\delta_v}\right).$$

So, for any starting vertex $u \neq v$,

$$Pr(A_t(v)|A_{t-1}(v)) = 1 - Pr(\overline{A_t(v)}|A_{t-1}(v))$$
$$= 1 - \sum_{y\in N_G(v)} Pr(X_{u,G}(t-1)=y|A_{t-1}(v))\frac{1}{d_G(y)}$$
$$\leq 1 - \left(\frac{\delta_v-1}{\delta_v} - O\left(\frac{\tau_0}{\delta_v}\right)\right)\sum_{y\in N_G(v)} Pr(X_{w,H}(\tau_0)=y)\frac{1}{d_G(y)}$$
$$= 1 - \left(\frac{\delta_v-1}{\delta_v} - O\left(\frac{\tau_0}{\delta_v}\right)\right)\sum_{y\in N_G(v)}\left(\frac{d_G(y)-1}{2|e(G)|-2d_G(v)} - O(n^{-3})\right)\frac{1}{d_G(y)}$$
$$= 1 - \left(\frac{\delta_v-1}{\delta_v} - O\left(\frac{\tau_0}{\delta_v}\right)\right)\frac{d_G(v)}{2|e(G)|}$$

where in the last equality we used Lemma 1. This concludes the proof, since by Bayes formula, $Pr(A_t(v)) = Pr(A_t(v)|A_{t-1}(v))\cdots Pr(A_{\tau_0+1}(v)|A_{\tau_0}(v))Pr(A_{\tau_0})$. $\square$

6. An upper bound on the cover time

Let G be an instance of $G_{n,m,p}$, where $m = n^\alpha$, $\alpha < 1$ and $p = 4\frac{\ln n}{m}$. Fix an arbitrary vertex u. Let $T_G(u)$ be the time that the random walk W_u on G needs to visit every vertex in $V(G)$. The following theorem shows that the cover time on G is optimal assuming that G is a "typical" instance of the $G_{n,m,p}$ model in this range, i.e. an instance that satisfies Lemmata 1, 2, 3 and has $\tau_0^{(H)} \leq \tau_0 = \Theta(\log n)$ for every $H = H(v)$ (the last assumption assures us that Lemma 5 can be applied). Note that almost every instance of $G_{n,m,p}$ in this range is "typical", since these requirements are satisfied whp.

Theorem 4. *The cover time C_u of the random walk starting from u is almost surely at most $\Theta(n\ln n)$.*

Proof. We will denote by U_t the number of vertices that have not been visited by W_u at step t. Clearly, the cover time of W_u satisfies

$$C_u = E[T_G(u)] = \sum_{t=0}^{\infty} Pr(T_G > t) = \sum_{t=0}^{\infty} Pr(U_t > 0) \leq \sum_{t=0}^{\infty}\min\{1, E[U_t]\}$$

by Markov's inequality. So, for any $t_0 > 0$,

$$C_u \leq t_0 + \sum_{t\geq t_0+1} E[U_t] = t_0 + \sum_{t\geq t_0+1}\sum_{v\in V(G)} Pr(A_t(v)) \tag{8}$$

where in the last equality we used the linearity of expectation and the fact that the events $\{v \in U_t\}$ and $A_t(v)$ are the same.

We set $t_0 = 5n\log n$. Since $\delta_v \geq n^{1-\alpha}\ln n$, for every $v \in V(G)$, by setting $Pr(A_{\tau_0}(v))$ equal to 1 in Lemma 5 and using the well known inequality $1 + x \leq e^x$, for any real x, we have that for all $t \geq t_0$,

$$Pr(A_t(v)) \leq \exp\left\{-\frac{td_G(v)}{2|e(G)|}\left(1 - O\left(\frac{\tau_0}{\delta_v}\right)\right)\right\} \leq \exp\left\{-(1-B)\frac{t|S_v|}{4n\ln n}\right\}$$

for some small constant $B > 0$. Note that for the final inequality we used the fact that $\tau_0 = o(\delta_v)$.

Now Eq. (8) becomes

$$\begin{aligned} C_u &\le 5n\ln n + \sum_{v\in V(G)}\sum_{t\ge t_0+1} \mathrm{e}^{(1-B)\frac{-t|S_v|}{4n\ln n}} \le 5n\ln n + 5n\ln n \sum_{v\in V(G)} \frac{1}{|S_v|}\mathrm{e}^{-\frac{5(1-B)}{4}|S_v|} \\ &\le 5n\ln n + 5n\ln n\left(\sum_{v:|S_v|\in M_1}\frac{1}{|S_v|}\mathrm{e}^{-|S_v|} + \sum_{v:|S_v|\in M_2}\frac{1}{|S_v|}\mathrm{e}^{-|S_v|}\right). \end{aligned} \tag{9}$$

Because of Lemma 1 and Lemma 3, the first sum is clearly $o(1)$ (just notice that for any v such that $|S_v| \in M_1$ we have that $|S_v| = \Theta(\ln n)$ and $D(|S_v|) \le \ln^3 n$). For the second sum, by Lemma 3 we have

$$\begin{aligned} \sum_{v:|S_v|\in M_2}\frac{1}{|S_v|}\mathrm{e}^{-\frac{|S_v|}{5}} &\le \sum_{k=1}^{m} D(k)\frac{1}{k}\mathrm{e}^{-k} \le \sum_{k=1}^{m} 2n\binom{m}{k}p^k(1-p)^{m-k}\frac{1}{k}\mathrm{e}^{-k} \\ &\le 7n\frac{1}{mp}\sum_{k=1}^{m}\binom{m+1}{k+1}p^{k+1}(1-p)^{m-k}\mathrm{e}^{-(k+1)} \\ &\le 7n\frac{1}{mp}\left(1-p+p\mathrm{e}^{-1}\right)^{m+1} = o(1). \end{aligned}$$

By (9) this means that $C_u \le \Theta(n\ln n)$ for any fixed vertex u. $\square$

Since the cover time $C = \max_{u\in V(G)} C_u$, and it is known that $C \ge (1-o(1))n\ln n$, we have proved

Theorem 5. *The cover time of an instance of $G_{n,m,p}$, with $m = n^{\alpha}$, $\alpha < 1$ and $p = 4\frac{\ln n}{m}$, is $C = \Theta(n\ln n)$ with high probability.*

7. Conclusions and future work

In this work, we investigated the expansion properties, the mixing time and the cover time of $G_{n,m,p}$ random intersection graphs for the non-trivial regime where $m = n^{\alpha}$, for $\alpha < 1$ and p very close to the connectivity threshold. We showed that the mixing time is logarithmic on the number of vertices and that the cover time is asymptotically optimal. Our analysis can be pushed further (although not without many technical difficulties) to provide even tighter results. However, the cover time and expansion properties in the case $\alpha = 1$ remain an open problem. It is worth investigating other important properties of $G_{n,m,p}$, such as dominating sets, existence of vertex disjoint paths between pairs of vertices etc.

Acknowledgements

We wish to warmly thank Noga Alon for pointing out to us the particular graph (two cliques of size $n/2$ with a perfect matching between them) that despite being a c-expander is not rapidly mixing, and also for providing material on his relevant research to us. We would also like to thank the anonymous reviewer who pointed out a gap in the analysis in Section 4.

References

[1] N. Alon, Personal communication, January 2007.
[2] N. Alon, J. Spencer, The Probabilistic Method, Second Edition, John Wiley & Sons, Inc., 2000.
[3] C. Avin, G. Ercal, On the cover time of random geometric graphs, in: Proceedings of the 32nd International Colloquium on Automata, Languages and Programmng, ICALP, 2005, pp. 677–689.
[4] B. Bollobás, Random Graphs, Second Edition, Cambridge University Press, 2001.
[5] C. Cooper, A. Frieze, The cover time of sparse random graphs, in: Random Structures and Algorithms, vol. 30, John Wiley & Sons, Inc., 2007, pp. 1–16.
[6] P. Diaconis, L. Saloff-Coste, Comparison theorems for reversible markov chains, Annals of Applied Probability 3 (3) (1993) 696–730.
[7] J. Díaz, M.D. Penrose, J. Petit, M. Serna, Approximating layout problems on random geometric graphs, Journal of Algorithms 39 (2001) 78–116.
[8] J. Díaz, J. Petit, M. Serna, Random geometric problems on $[0, 1]^2$, in: J. Rolim, M. Luby, M. Serna (Eds.), Randomization and Approximation Techniques in Computer Science, in: Lecture Notes in Computer Science, vol. 1518, Springer Verlag, Berlin, 1998, pp. 294–306.
[9] J. Díaz, J. Petit, M. Serna, A random graph model for optical networks of sensors, in: 1st International Workshop on Efficient and Experimental Algorithms, WEA, 2003. Also in the IEEE Transactions on Mobile Computing Journal 2 (3) (2003) 186–196.
[10] C. Efthymiou, P.G. Spirakis, On the existence of hamilton cycles in random intersection graphs, in: Proceedings of the 32nd International Colloquium on Automata, Languages and Programmng, ICALP, 2005, pp. 690–701.
[11] J.A. Fill, E.R. Sheinerman, K.B. Singer-Cohen, Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $G(n, m, p)$ and $G(n, p)$ models, Random Structures and Algorithms 16 (2) (2000) 156–176.
[12] E. Godehardt, J. Jaworski, Two models of random intersection graphs for classification, in: O. Opitz, M. Schwaiger (Eds.), Studies in Classification, Data Analysis and Knowledge Organisation, Springer Verlag, Berlin, Heidelberg, New York, 2002, pp. 67–82.
[13] M. Jerrum, A. Sinclair, Approximate counting, uniform generation and rapidly mixing markov chains, Information and Computation 82 (1989) 93–133.
[14] J. Jonasson, On the cover time of random walks on random graphs, Combinatorics, Probability and Computing 7 (1998) 265–279.
[15] M. Karoński, E.R. Scheinerman, K.B. Singer-Cohen, On random intersection graphs: The subgraph problem, Combinatorics, Probability and Computing Journal 8 (1999) 131–159.
[16] E. Marczewski, Sur deux propriétés des classes d' ensembles, Fund. Math. 33 (1945) 303–307.

[17] S. Nikoletseas, K. Palem, P.G. Spirakis, M. Yung, Short vertex disjoint paths and multiconnectivity in random graphs: Reliable network computing, in: Proceedings of the 21st International Colloquium on Automata, Languages and Programming, ICALP, in: Lecture Notes in Computer Science, vol. 820, Springer Verlag, 1994, pp. 508–519. Also, in the Special Issue on Randomized Computing of the International Journal of Foundations of Computer Science, IJFCS, vol. 11, No. 2, World Scientific Publishing Company, 2000, pp. 247–262.
[18] S. Nikoletseas, C. Raptopoulos, P.G. Spirakis, The existence and efficient construction of large independent sets in general random intersection graphs, in: Proceedings of the 31st International Colloquium on Automata, Languages and Programming, ICALP, in: Lecture Notes in Computer Science, Springer Verlag, 2004, pp. 1029–1040. Also, in the Theoretical Computer Science (TCS) 406 (3) (2008) 215–224 (Special issue on Algorithmic Aspects of Global Computing).
[19] S. Nikoletseas, C. Raptopoulos, P.G. Spirakis, Expander properties and the cover time of symmetric random intersection graphs, in: Proceedings of the 32nd International Symposium on Mathematical Foundations of Computer Science, I 2007, 2007.
[20] S. Nikoletseas, C. Raptopoulos, P.G. Spirakis, The second eigenvalue of random walks on symmetric random intersection graphs, in: Proceedings of the 2nd International Conference on Algebraic Informatics, CAI 2007, 2007.
[21] S. Nikoletseas, P.G. Spirakis, Expander properties in random regular graphs with edge faults, in: Proceedings of the 12th Annual Symposium on Theoretical Aspects of Computer Science, STACS, in: Lecture Notes in Computer Science, vol. 900, Springer Verlag, 1995, pp. 421–432.
[22] M. Penrose, Random Geometric Graphs, in: Oxford Studies in Probability, 2003.
[23] C. Raptopoulos, P.G. Spirakis, Simple and efficient greedy algorithms for hamilton cycles in random intersection graphs, in: Proceedings of the 16th International Symposium on Algorithms and Computation, ISAAC, 2005, pp. 493–504.
[24] A. Sinclair, Algorithms for random generation and counting: A Markov chain approach, PhD Thesis, University of Edimburg, 1988.
[25] A. Sinclair, Algorithms for Random Generation and Counting, Birkhauser, 1992.
[26] K.B. Singer-Cohen, Random intersection graphs, PhD thesis, John Hopkins University, 1995.
[27] D. Stark, The vertex degree distribution of random intersection graphs, Random Structures & Algorithms 24 (3) (2004) 249–258.